

Author(s): Oliver Dibben and Andrew J. Easton

Article Title: Mutational analysis of the gene start sequences of pneumonia virus of mice

Year of publication: 2007

Link to published article:
http://dx.doi.org/10.1016/j.virusres.2007.06.009

# Mutational analysis of the gene start sequences of pneumonia virus of mice

**Oliver Dibben[1] and Andrew J. Easton***
**Department of Biological Sciences**
**University of Warwick**
**Coventry**
**CV4 7AL**
**UK**

* Corresponding author:
Tel: +44 (0)2476 524727
Fax: +44 (0)2476 523568
E-mail:a.j.easton@warwick.ac.uk

[1] Current Address:

University of Oxford,
The Jenner Institute,
Compton, Newbury
Berkshire,
RG20 7NN

E-mail:Oliver.Dibben@jenner.ac.uk]

Abstract: 200 words

Main text: 3268 words

Key words: Pneumonia virus of mice, pneumoviruses, gene start sequence, transcription

**Abstract**

The transcriptional start sequence of pneumonia virus of mice is more variable than that of the other pneumoviruses, with five different nine-base gene start (GS) sequences found in the PVM genome. The sequence requirements of the PVM gene start signal, and the efficiency of transcriptional initiation of the different virus genes, was investigated using a reverse genetics approach with a minigenome construct containing two reporter genes. A series of GS mutants were created, where each of the nine bases of the gene start consensus sequence of a reporter gene was changed to every other possible base, and the resulting effect on initiation of transcription was assayed. Nucleotide positions 1, 2 and 7 were found to be most sensitive to mutation whilst positions 4, 5 and 9 were relatively insensitive. The L gene GS sequence was found to have only 20% of the activity of the consensus sequence whilst the published M2 gene start sequence was found to be non-functional. A minigenome construct in which the two reporter genes were separated by the F-M2 gene junction of PVM was used to confirm the presence of two alternative, functional, GS sequences that could both drive the transcription of the PVM M2 gene.

Pneumonia virus of mice (PVM) is a member of the family *Paramyxoviridae*, subfamily *Pneumovirinae* and is in the Pneumovirus genus. PVM is closely related to human respiratory syncytial virus (RSV) and is currently being developed as an animal model for RSV (Bonville et al., 2003; Cook et al., 1998; Domachowske et al., 2000a; Domachowske et al., 2002; Domachowske et al., 2000b; Thorpe and Easton, 2005). PVM has a single stranded, non-segmented, negative sense RNA (NNS-RNA) genome of between 14,885 (Strain J3666) and 14,887 (Strain 15) nucleotides in length (Thorpe and Easton, 2005: Accession numbers AY743909 and AY743910). The genome RNA encodes ten genes as individual transcription units and generates eleven major protein products (Ahmadian et al., 1999; Chambers et al., 1990a; Chambers et al., 1990b; Thorpe and Easton, 2005). The gene order and the major proteins expressed during PVM infection are similar to those in RSV, the main distinction being that the M2 and L genes of RSV overlap by 68 nucleotides whereas the M2 and L genes of PVM are separated by a distinct intergenic region (Chambers et al., 1990b; Collins et al., 1987). The presence of the overlap in RSV results in premature termination of L gene mRNA at the M2 gene end so reducing the level of full length L gene mRNA (Collins et al., 1987).

NNS-RNA virus RNA synthesis is carried out by the viral RNA-dependent RNA polymerase, which always initially binds the leader region at the 3' terminus of the genome. The minimal competent unit for transcription in NNS-RNA viruses is the ribonucleoprotein complex, consisting of viral genomic RNA complexed with the nucleoprotein (N), the phosphoprotein (P) and the large polymerase protein (L). In pneumoviruses an additional protein, the M2-1 protein, acts as a transcription elongation factor (Collins et al., 1996; Grosfeld et al., 1995). Transcription of pneumovirus genes is controlled by initiation at a nine nucleotide consensus transcription initiation sequence called a gene start (GS), with termination at a consensus 13-16 nucleotide transcription termination sequence called a gene

end (GE). Transcription progresses in a sequential interrupted fashion along the length of the genome whereby termination of transcription at the GE leads to either dissociation of the polymerase from the template or re-initiation of transcription at the subsequent GS sequence (Collins et al., 1986). This results in the abundance of mRNA species being controlled by both proximity of a gene to the 3' terminus of the genome, and efficiency of re-initiation of transcription at the gene start sequence (Dickens et al., 1984; Iverson and Rose, 1981; Pringle and Easton, 1997). In recent years it has become apparent that the initiation of transcription at GS sequences may depend on the precise nature of the sequence. Both RSV and avian pneumovirus (APV) possess a highly conserved GS sequence, only varying in sequence in the GS which directs transcription of the L gene (Chambers et al., 1990b; Collins et al., 1986; Li et al., 1996; Ling et al., 1995; Ling et al., 1992; Randhawa et al., 1996; Yu et al., 1991; Yu et al., 1992). In contrast, the PVM GS has one of five possible sequences (Table 1), the consensus sequence being AGGAyAArT, suggesting the possibility of differential expression of genes or a tolerance of variation in GS sequence (Chambers et al., 1990a; Chambers et al., 1990b). The GS sequences of both APV and RSV have been subjected to mutational analysis that identified residues important for the initiation of transcription (Edworthy and Easton, 2005; Kuo et al., 1997). Here, we describe a similar analysis of the role of each of the nine individual bases of the PVM GS sequence using a reverse genetics approach.

To assess the efficiency of initiation of transcription by different mutants of the PVM GS signal a dicistronic minigenome was constructed (Figure 1), containing two genes, similar to those described previously for APV and RSV (Edworthy and Easton, 2005; Kuo et al., 1996). The first gene in the PVM minigenome encoded bacterial chloramphenicol acetyl transferase (CAT) and was flanked by the N gene GS signal (AGGATAAAT) and the L gene GE signal (TAGTTAACAAAAAA). The second gene encoded enhanced green fluorescent protein (EGFP) and was also flanked by the N gene GS sequence and the L gene GE

sequence. The N gene GS sequence was used as it most closely resembled the consensus GS sequence. The intergenic region separating the two genes was that of the N-P gene junction, modified to contain a *Bgl*II site for cloning purposes (TGAGATCTGT (*Bgl*II site underlined)). For the construction, two monocistronic minigenomes were constructed, one containing the CAT gene (pKozmg) and one the EGFP gene (pEGFPmg: Figure 1), these were then fused together to produce a dicistronic construct. The second cistron of the dicistronic construct pCAT-EGFPmg was synthesised by PCR from the EGFP minigenome pEGFPmg. Primers CATEGFP 1 (all primers used in the construction of the various minigenomes are shown in Table 2) and CATEGFP 3 were used to amplify the EGFP gene along with, at the 3' end, the PVM N gene GS sequence with the addition of an intergenic region, and at the 5' end the L gene GE sequence. This second cistron was inserted into the CAT minigenome pKozmg, downstream of the CAT gene to create a dicistronic construct. Every position of the EGFP reporter gene GS signal was mutated to each of the three other possible bases using either quickchange mutagenesis (Stratagene) or standard two-step PCR mutagenesis, as described previously (Edworthy and Easton, 2005), resulting in a set of 28 mutants plus the N GS control construct. Mutagenic primers of 48 nucleotides in length were used for both techniques (all mutagenic primer sequences are available from authors on request). For two-step PCR mutagenesis, flanking primers GSA, containing an *Nco*I restriction site, and GSB, containing a *Not*I restriction site, were used to produce an insert with a mutated GS sequence that could be ligated into the parental plasmid pCAT-EGFPmg.

The dicistronic minigenome construct was transfected into BSRT-7 cells (Buchholz et al., 1999) using Fugene6 transfection reagent (Roche) along with plasmids carrying the N, P, L and M2-1 genes of PVM Strain 15 (Thorpe and Easton, 2005). The plasmids were transfected in the ratios (per $10^6$ cells), 0.4μg minigenome, 0.4μg N gene plasmid, 0.2μg P gene plasmid, 0.2μg L gene plasmid and 0.1μg M2-1 ORF plasmid. The effect on initiation

of transcription from the mutant GS sequences was monitored by measuring changes in levels of EGFP protein expression. Both CAT (Marriott et al., 2001) and EGFP protein levels were assayed by ELISA, and the level of EGFP protein expressed for each mutant was calculated as a percentage of that seen with the intact N GS sequence. The data were then normalised using levels of CAT protein expression derived from the non-mutated first cistron as described previously (Edworthy and Easton, 2005).

Relative levels of EGFP protein expression derived from each mutant are shown in figure 2. All mutations made to the GS signal resulted in a decrease in expression of EGFP protein, with the level of the decrease varying greatly between different mutants. Certain positions in the PVM GS signal were particularly tolerant or intolerant to mutation. As can be seen, positions 1, 2 and 7 of the PVM GS sequence were intolerant to mutation as any nucleotide substitution at these positions resulted in almost complete loss of EGFP protein expression, suggesting that mutation in these locations results in the abrogation of binding of the viral polymerase complex to the GS sequence. In contrast, positions 4, 5 and 9 of the PVM GS signal appeared to be more tolerant of mutation, with most nucleotide changes at these positions resulting in retention of more than of 20% of EGFP protein expression. In the APV and RSV GS sequences, position 7 has previously been found to be highly sensitive to mutation, suggesting that this is an essential component in the initiation of transcription in all pneumoviruses. As for PVM, position 5 has been shown to be more tolerant to mutation as in APV and RSV indicating that this is a common feature of the pneumovirus GS sequence (Edworthy and Easton, 2005; Kuo et al., 1997)

Positions 3 and 8 of the PVM GS sequence displayed a preference for purines. At position 3, substituting the conserved guanine residue for a pyrimidine resulted in the loss of almost all EGFP protein expression, while replacing it with adenine gave 40% of maximal expression. Similarly, at position 8 substitution of conserved adenine for guanine generated

60% of maximal EGFP protein levels whereas substitution of a pyrimidine resulted in much greater decrease of EGFP protein expression. This preference for either pyrimidine or purine at specific locations has also been observed for both RSV and APV. Position 8 in both PVM and RSV GS sequences display a similar pattern of sensitivity to mutation, with substitution of a purine for a purine (PVM) or a pyrimidine for a pyrimidine (RSV), resulting in conservation of activity, while any other change abolishes activity. Position 9 of the PVM GS signal was relatively tolerant to sequence change, with EGFP protein expression levels of between 22% and 56% of maximal levels for all mutations, whereas position 9 of both APV and RSV GS signals have been shown to be very sensitive to mutation (Edworthy and Easton, 2005; Kuo et al., 1997).

To confirm that the changes in protein expression observed were due to changes in the efficiency of initiation of transcription, rather than an independent effect on translation, the levels of mRNA synthesised from both reporter genes by a selection of mutants were analysed (Figure 3). A series of mutants were chosen to represent those affected either only slightly, moderately or severely by the mutations. Transfections were conducted as above but 48 hours post-transfection, mRNA was purified from the cells using a Sigma GenElute™ Direct mRNA Miniprep Kit. Purified mRNA was separated on a glyoxal-agarose gel and detected by northern blotting using non-radioactive DIG-labelled riboprobes specific for both CAT and EGFP (Marriott et al., 1999). CAT mRNA levels remained constant for the various mutants examined, as expected. However, the levels of EGFP mRNA detected varied in conjunction with changes in EGFP protein expression. These data confirmed that the changes in EGFP protein expression were due to variations in the efficiency of transcription as a result of the mutations introduced into the GS sequence. Uniform levels of read-through mRNA confirmed that alterations to the EGFP GS signal were not affecting the processivity of the polymerase across the gene junction.

The series of mutants described here includes all GS sequences present in the PVM genome (Table 1). From comparison of these sequences with the mutants analysed several observations can be made. Firstly, the GS sequence used by the NS1 and NS2 genes contains two deviations from the N GS sequence (U5C and A8G), resulting in 45% lower expression. The individual mutants U5C and A8G result in a similar drop in protein expression (Figure 2), showing that the two differences in sequence are not cumulative. Secondly, eight of the ten GS sequences found in the PVM genome are either identical to the N GS sequence used here or contain the changes U5C or A8G, both of which are still relatively efficiently transcribed. Thus, although the PVM GS sequence is more variable than those of RSV and APV, the variations do not have a dramatic effect on the efficiency of transcription of virus genes. This indicates that the majority of the variability seen in the PVM GS sequence is tolerated by the virus polymerase and does not provide a significant method of regulation of virus genes. However, due to the sequential transcription of genes the cumulative effect of these moderate decreases in transcription across several gene junctions could result in significant down-regulation of transcription from promoter-distal genes. The other two GS sequences occurring naturally in the PVM genome are the L gene (A6C compared to the N GS) and the M2 gene (A6G and A8G compared to the N GS: Table 1) GS signals. The L gene GS of PVM is only 20% as active as the N gene GS (indicated by an asterisk in Figure 2). This contrasts with the APV and RSV L gene GS sequences which were more divergent in sequence from the consensus (each containing three nucleotide differences) but which resulted in protein expression levels of 50% and 100%, respectively (Edworthy and Easton, 2005; Kuo et al., 1997). This suggests that attenuation of L gene expression due to a decreased efficiency of initiation of transcription is a more significant mode of regulation in PVM than either APV or RSV. However, transcription of the RSV L gene is severely attenuated due to the presence of the M2/L gene overlap (Collins et al., 1987) and this results

in a level of expression of L mRNA much lower than would be predicted from the efficiency of the gene start sequence. Surprisingly, based on the EGFP protein expression data reported here, the proposed PVM M2 GS sequence is non-functional, giving EGFP protein levels of only 0.5% of that seen with the N gene start sequence. The A6G sequence contained in the putative M2 gene GS sequence, on its own, results in complete shutoff of EGFP protein expression. The original assignment of the M2 gene start sequence was based on sequences derived from a cDNA library (Chambers et al., 1990b). However, it had been previously noted that the 5' end sequences in cDNA clones of PVM mRNAs frequently contained an inversion due to the presence of inverted repeats (Barr et al., 1991) and this may have led to a misidentification of the 5' end of the M2 mRNA. Given the potential role of the M2 gene products in pneumovirus gene expression and the likely consequential effect on L gene expression a significant down regulation of PVM M2 gene transcription of this order appeared unlikely. These data therefore suggested that the M2 GS sequence of PVM may have been originally assigned incorrectly and that the real M2 GS signal is more active. Consideration of the PVM genome sequence identified two sequences upstream of the designated M2 GS signal, and within the sequence previously identified as the intergenic region between the F and M2 genes, that fit the consensus sequence and may therefore act as GS signals (Figure 4A). These have the sequences AGGAUAAGU (-53 bases, referred to here as GS1) and AGGAUAAGG (-10 bases upstream, referred to here as GS2). Either of these sequences could potentially function as a GS signal, the GS1 sequence having only an A8G substitution and the GS2 sequence containing the sequence changes A8G and U9G, neither of which are overly detrimental to transcription while, as indicated in Table 1 the previously identified M2 GS sequence, contains the alterations A6G and U8G and is designated as GS3 below.

To investigate the contribution of each of these sequences to the initiation of transcription of the PVM M2 gene a dicistronic minigenome construct containing the F-M2 intergenic region, designated pF-M2GS, was used. Transcription of the EGFP gene in this construct was therefore under the control of all three putative M2 GS sequences. The F-M2 IG region was amplified using primers CATEGFP-M2GSA and CATEGFP-M2GSD (Table 2). The EGFP gene of pCAT-EGFPmg was amplified using primers CATEGFP-M2GSC and CATEGFP 3 and these DNA fragments were fused together by overlap PCR using primers CATEGFP-M2GSA and CATEGFP 3. The resulting product was inserted into pCAT-EGFPmg and the L gene GE sequence of the CAT gene was corrected to an F GE sequence by quickchange mutagenesis, using primers L-FGE+ and L-FGE-. This plasmid, designated pF-M2IG, contained three GS sequences, GS1, GS2 and GS3. The role of each alternate GS sequence in transcription was assessed as described for the other GS mutants, only the levels of EGFP protein expression were calculated as a percentage of that given with all three GS sequences intact.

The F-M2 intergenic region gave levels of EGFP protein expression of almost 10% of that given by the N GS positive control (Figure 2). This confirmed that the intergenic region containing all three M2 GS sequences could drive expression of EGFP protein and was significantly more efficient at initiating transcription than GS3 alone. For the purpose of this analysis GS3 was considered to be non-functional as it gave levels of EGFP protein expression only marginally above background (Figure 2). Given that the first gene start sequence (GS1) in the intergenic region contained only the mutation A8G which had previously been shown to reduce expression to 60% of that seen with the N gene start the observation that the entire intergenic region gave only 10% activity was unexpected. Two possible explanations for this discrepancy are possible. Firstly, there may be less efficient termination of transcription at the F gene end sequence compared to the level with the L gene

end sequence leading to different levels of reinitiation at the subsequent gene start. The efficiency of termination at different gene end sequences has not been assessed. Secondly, the lengths of the intergenic regions in the two constructs used are not identical and this may affect reinitiation frequency. The first two sequences, GS1 and GS2, were individually mutated to non-functional GS sequences to assess the role of each in the initiation of transcription of the EGFP gene. This was achieved by the introduction of an A→G nucleotide change at residue 1 of the GS sequence by quickchange mutagenesis, with GS1 mutated using primers GS1-A1G+ and GS1-A1G- and GS2 mutated using primers GS2-A1G+ and GS2-A1G-. As shown in Figures 2 and 3 the alteration of the first A residue of the consensus GS sequence completely ablated transcriptional activity. The effect of the deletion of either GS1 or GS2 in the F-M2 intergenic region on EGFP protein expression is shown in Figure 4B. The construct containing all three GS sequences is labeled M2GS-123. The construct with the mutated GS1 is labeled M2GS-23, and the construct with the mutated GS2 is labeled M2GS-13. It can be seen that mutation of either GS1 or GS2 had a deleterious effect on the expression of EGFP protein. In the absence of GS1 the level of EGFP protein expression dropped to approximately 25% of that seen with all three GS sequences. In the absence of GS2 the decrease in EGFP protein expression was less dramatic, falling to approximately 60% of that with all three GS sequences. This suggests that both GS1 and GS2 can initiate transcription but that GS1 is more important than GS2. This agrees with the results obtained from the mutational analysis described above with the GS1 sequence, containing only an A8G nucleotide substitution when compared to the consensus GS sequence, expected to more efficiently initiate transcription than the GS2 sequence, which contained both A8G and U9G nucleotide substitutions. Overall, these data suggest that, the originally designated GS sequence of the M2 gene of PVM is not involved in transcription of

the M2 gene, and that two other upstream GS sequences could both contribute to the transcription of the M2 gene, with the first GS sequence being more important.

In conclusion mutational analysis of the gene start sequence of PVM has shown that it, like the other pneumoviruses, is variable in its sensitivity to mutation. Positions 1, 2 and 7 of the nine base sequence are highly intolerant to mutation, suggesting that these bases play an essential role in interaction with the viral polymerase complex during initiation of transcription. Position 7 is also particularly sensitive to mutation in both APV and RSV, suggesting that this residue plays a key role in initiation of transcription across the pneumovirus sub-family. Positions 4, 5 and 9 of the PVM GS sequence were found to be more tolerant to mutation, and of these, position 5 has been shown previously to be less sensitive to mutation in both APV and RSV (Edworthy and Easton, 2005; Kuo et al., 1997). Position nine of the PVM GS is tolerant to mutation, as is position 10 of the RSV GS (Kuo et al., 1997), and this contrasts with APV where the terminal base of the GS was shown to be sensitive to mutation (Edworthy and Easton, 2005). This identifies a possible distinction between the different genera of the *Pneumovirinae*, the 5' nucleotides of the GS being less important in polymerase binding in pneumoviruses than in metapneumoviruses. Comparing the figures of relative protein expression levels produced for all three viruses, it can be seen that the most sensitive areas of the pneumovirus GS sequences cluster around bases 1, 2, 6 and 7, identifying these regions as being important in the initiation of transcription, with the central region of the sequence seemingly playing less of a role.

The M2 gene GS sequence of PVM was identified as having only minimal activity in the initiation of transcription. Two alternative sequences were identified upstream of the designated M2 GS sequence which could potentially act as GS signals (designated GS1 and GS2). These factors suggested that the GS sequence of the PVM M2 gene was originally assigned incorrectly. The first GS sequence (GS1) following the F GE sequence appears to

be most important. This fits with the proposed mechanism for transcription where following termination of transcription the polymerase complex moves along the genome template until it encounters the next available GS sequence at which it can initiate transcription of the next gene.

**Acknowledgements.** O.D. was supported by a postgraduate studentship awarded by the Biotechnology and Biological Sciences Research Council.

**References**


Ahmadian, G., Chambers, P. and Easton, A.J. (1999) Detection and characterisation of proteins encoded by the second ORF of the M2 gene of pneumoviruses. J. Gen. Virol. 80, 2011-2016.

Barr, J., Chambers, P., Pringle, C.R. and Easton, A.J. (1991). Sequence of the major nucleocapsid protein gene of pneumonia virus of mice: sequence comparisons suggest structural homology between nucleocapsid proteins of pneumoviruses, paramyxoviruses, rhabdoviruses and filoviruses. J. Gen. Virol. 72, 677-685.

Bonville, C.A., Easton, A.J., Rosenberg, H.F. and Domachowske, J.B. (2003) Altered pathogenesis of severe pneumovirus infection in response to combined antiviral and specific immunomodulatory agents. J Virol. 77(2), 1237-44.

Buchholz, U.J., Finke, S. and Conzelmann, K.-K. (1999) Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J. Virol. 73, 251-259.

Chambers, P., Barr, J., Pringle, C.R. and Easton, A.J. (1990a) Molecular cloning of pneumonia virus of mice. J. Virol.64, 1869-1872.

Chambers, P., Matthews, D.A., Pringle, C.R. and Easton, A.J. (1990b) The nucleotide sequences of intergenic regions between nine genes of pneumonia virus of mice establish the physical order of these genes in the viral genome. Vir. Res. 18, 263-270.

Collins, P.L., Dickens, L.E., Buckler-White, A., Olmsted, R.A., Spriggs, M.K., Camargo, E. and Coelingh, K.L.v.W. (1986) Nucleotide sequences from the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. Proc.Natl. Acad. Sci. USA 83, 4594-8.

Collins, P.L., Hill, M.G., Cristina, J. and Grosfeld, H. (1996) Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. Proc.Natl. Acad. Sci. USA 93, 81-85.

Collins, P.L., Olmsted, R.A., Spriggs, M.K., Johnson, P.R. and Buckler-White, A.J. (1987) Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. Proc.Natl. Acad. Sci. USA 84, 5134-8.

Cook, P.M., Eglin, R.P. and Easton, A.J. (1998) Pathogenesis of pneumovirus infections in mice : detection of pneumonia virus of mice and human respiratory syncytial virus mRNA in lungs of infected mice by in-situ hybridisation. J. Gen. Virol. 79, 2411-2417.

Dickens, L.E., Collins, P.L. and Wertz, G.W. (1984) Transcriptional mapping of human respiratory syncytial virus. J. Virol. 52, 364-9.

Domachowske, J.B., Bonville, C.A., Dyer, K.D., Easton, A.J. and Rosenberg, H.F. (2000a) Pulmonary eosinophilia and production of MIP-1a are prominent responses to infection with pneumonia virus of mice. Cell. Immunol. 200, 98-104.

Domachowske, J.B., Bonville, C.A., Easton, A.J. and Rosenberg, H.F. (2002) Differential expression of proinflammatory cytokine genes in vivo in response to pathogenic and nonpathogenic pneumovirus infections. J Infect Dis 186(1), 8-14.

Domachowske, J.B., Bonville, C.A., Gao, J., Murphy, P.M., Easton, A.J. and Rosenberg, H.R. (2000b) The chemokine macrophage-inflammatory protein-1a and its receptor CCR1 control pulmonary inflammation and antiviral host defense in paramyxovirus infection. J.Immunol. 165, 2677-2682.

Edworthy, N.L. and Easton, A.J. (2005) Mutational analysis of the avian pneumovirus conserved transcriptional gene start sequence identifying critical residues. J. Gen. Virol. 86, 3343-7.

Grosfeld, H., Hill, M.G. and Collins, P.L. (1995) RNA replication by respiratory syncytial virus (RSV) is directed by the N, P and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. J. Virol. 69, 5677-5686.

Iverson, L.E. and Rose, T.K. (1981) Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription. Cell 23, 477-84.

Kuo, L., Fearns, R. and Collins, P.L. (1997) Analysis of the gene start and gene end signals of human respiratory syncytial virus: quasi-templated initiation at position 1 of the encoded mRNA. J. Virol. 71, 4944-4953.

Kuo, L., Grosfeld, H., Cristina, J., Hill, M.G. and Collins, P.L. (1996) Effect of mutations in the gene-start and gene-end sequence motifs on transcription of monocistronic and dicistronic minigenomes of respiratory syncytial virus. J. Virol. 70(10), 6892-6901.

Li, J., Ling, R., Randhawa, J.S., Shaw, K., Davis, P.J., Juhasz, K., Pringle, C.R., Easton, A.J. and Cavanagh, D. (1996) Sequence of the nucleocapsid protein gene of subgroup-A and subgroup-B avian pneumoviruses. Vir. Res. 41(2), 185-191.

Ling, R., Davis, P.J., Yu, Q.Z., Wood, C.M., Pringle, C.R., Cavanagh, D. and Easton, A.J. (1995) Sequence and in-vitro expression of the phosphoprotein gene of avian pneumovirus. Vir. Res. 36, 2-3.

Ling, R., Easton, A.J. and Pringle, C.R. (1992) Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene order different from that of other pneumoviruses. J. Gen.Virol. 73, 1709-15.

Marriott, A.C., Smith, J.M. and Easton, A.J. (2001) Fidelity of leader and trailer sequence usage by the respiratory syncytial virus and avian pneumovirus replication complexes. J. Virol. 75, 6265-6272.

Marriott, A.C., Wilson, S.D., Randhawa, J.S. and Easton, A.J. (1999) A single amino acid substitution in the phosphoprotein of respiratory syncytial virus confers thermosensitivity in a reconstituted RNA polymerase system. J. Virol. 73, 5162-5165.

Pringle, C.R. and Easton, A.J. (1997) Monopartite negative strand RNA genomes. Seminars in Virology 8, 49-57.

Randhawa, J.S., Pringle, C.R. and Easton, A.J. (1996) Nucleotide-sequence of the matrix protein gene of a subgroup-B avian pneumovirus. Vir. Genes 12(2), 179-183.

Thorpe, L.C. and Easton, A.J. (2005) Genome sequence of the non-pathogenic strain 15 of pneumonia virus of mice and comparison with the genome of the pathogenic strain J3666. J. Gen. Virol. 86, 159-169.

Yu, Q., Davis, P.J., Barrett, T., Binns, M.M., Boursnell, M.E.G. and Cavanagh, D. (1991) Deduced amino acid sequence of the fusion glycoprotein of turkey rhinotracheitis virus has greater identity with that of human respiratory syncytial virus, a pneumovirus, than that of paramyxoviruses and morbilliviruses. Journal of General Virology 72, 75-81.

Yu, Q., Davis, P.J., Brown, T.D.K. and Cavanagh, D. (1992) Sequence and in vitro expression of the M2 gene of turkey rhinotracheitis pneumovirus. J.Gen. Virol. 73, 1355-63.

## FIGURE LEGENDS

**Table 1**. The gene start sequences of the 10 genes of PVM are more variable than those of other pneumoviruses. The nine residue sequence is conserved across six of the nine positions but is variable at positions 5, 6 and 8. Conserved residues are shaded and the consensus gene start sequences is shown.

**Table 2**. Sequences of oligonucleotide primers used in the construction of PVM minigenomes.

**Figure 1.** Diagram of construction of the PVM dicistronic minigenome. The DNA minigenome construct contained two reporter genes, CAT and EGFP, as separate transcription units derived from separate monocistronic minigenome constructs, and produced a genome sense RNA from T7 transcription. Both genes were under the control of PVM consensus gene start (GS) and gene end (GE) signals and were flanked by the PVM leader and trailer regions. GS marked with * was mutated to allow assessment of effect of mutations on the transcription of the EGFP gene. T7 transcription gives rise to a negative sense minigenome RNA with termini defined by the T7 promoter (5') and HDV ribozyme (3'). Plasmid encoded RNP components encapsidate, transcribe and replicate the minigenome RNA. RNA synthesis always initiates at the leader region of the negative sense RNA and results in synthesis of either mRNA, defined by the GS and GE sequences, or full length anti-minigenome RNA.

**Figure 2.** Figure showing relative expression of EGFP protein for all GS mutants. The wild-type (N gene) GS sequence is displayed below the x axis (AGGAUAAAU), while nucleotide changes at each position are shown on the x axis. Levels of EGFP protein expression are displayed as percentage of that produced by the consensus GS sequence (NGS – black bar), and are averages of three independent experiments.

Gene start sequences present in the PVM genome containing more than one nucleotide change from the wild-type are shown with white bars (NS1/NS2, M2). Bar labelled with * is the mutation corresponding to the L gene GS sequence. Results obtained using a construct containing the PVM F-M2 gene junction are shown with diagonal shading (designated F-M2 IG).

**Figure 3.** Comparison of relative EGFP protein expression with EGFP mRNA levels for a selection of mutants. Northern blot analysis was used to detect both CAT mRNA and EGFP mRNA produced by each mutant construct, with CAT mRNA acting as an internal control for initiation of transcription from the leader sequence.

**Figure 4.** Figure showing the sequences of, and relative EGFP protein expression from, three putative GS sequences of the PVM M2 gene.

**A**. Sequence of the PVM F-M2 intergenic region. Sequences are shown as positive sense RNA. The gene end of the F gene is shown in bold and the three putative M2 gene start sequences GS1, GS2 and GS3 and underlined.

**B**. The efficiency of EGFP protein expression with either GS1 or GS2 deleted (M2GS-23 and M2GS-13, respectively) is shown as a percentage of that given with all three GS sequences (M2GS-123), shown as 100%.

**Table 1.**

| GENE | GENE START SEQUENCE |
|---|---|
| **N, P and SH** | **AGGAUAAAU** |
| **NS1 and NS2** | **AGGACAAGU** |
| **M and F** | **AGGACAAAU** |
| **G** | **AGGAUAAGU** |
| **M2** | **AGGAUGAGU** |
| **L** | **AGGAUCAAU** |
| **Consensus** | **AGGA U a AaA** |

**Table 2.**

| Oligonucleotide name | Sequence |
|---|---|
| CATEGFP 1 | CGATGTTAACAAAAAACTGAGATCTGTAGGATAAATACATA TGGTGAGCAAG |
| CATEGFP 3 | GTGAGATATCAATTTTTTGTTAACTATA |
| GSA | CCCGTTTTCACCATGGGCAAATAT |
| GSB | TAGAGTCGCGGCCGCTTTACTTGT |
| CATEGFP-M2GSA | GCCGTAGTTAACAAAAAACTTAGGATAAGTGACAATCCAG |
| CATEGFP-M2GSD | CTCGCCCTTGCTCACACTCATCCTACCTTATCCTTGA |
| CATEGFP-M2GSC | TAAGGTAGGATGAGTGTGAGCAAGGGCGAGGAGCTGT |
| L-FGE+ | ATTATTTTAATTTGTCTATAGTTAATTAAAAACTTAGGATAA GTGACAATCC |
| L-FGE- | GGATTGTCACTTATCCTAAGTTTTTAATTAACTATAGACAAA TTAAAATAAT |
| GS1-A1G+ | TTTGTCTATAGTTAATTAAAAACTTGGGATAAGTGACAATCC AGACCCAAC |
| GS1-A1G- | GTTGGGTCTGGATTGTCACTTATCCCAAGTTTTTAATTAACTA TAGACAAA |
| GS2-A1G+ | GACCCAACACCTCTTTCAACTCTCAGGGATAAGGTAGGATGA GTGTGAGCA |
| GS2-A1G- | TGCTCACACTCATCCTACCTTATCCCTGAGAGTTGAAAGAGGT GTTGGGTC |